Version 4, Paris 11 December 2009
Patric.Amcoff@oecd.org

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

DRAFT PROPOSAL FOR A NEW GUIDELINE

In Vitro Skin Irritation: Reconstructed Human *Epidermis* (RhE) Test Method

INTRODUCTION

1. Skin irritation refers to the production of reversible damage to the skin following the application of a test chemical for up to 4 hours [as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS)](1). This Test Guideline provides an *in vitro* procedure that may be used for the hazard identification of irritant chemicals (substances and mixtures) in accordance with UN GHS and EU CLP Category 2 (1)(2)(3), as well as non-classified chemicals, *i.e.* UN GHS No-Category, for member countries or regions that do not adopt the optional UN GHS Category 3 (mild irritants). Depending on the regulatory framework and the classification system in use, this test method may be used to determine the skin irritancy of chemicals as a stand-alone replacement test for *in vivo* skin irritation testing, or as a partial replacement test, within a tiered testing strategy (4).

2. The assessment of skin irritation has typically involved the use of laboratory animals [OECD Test Guideline 404; adopted in 1981 and revised in 1992 and 2002](4). In relation to animal welfare concerns, TG 404 was revised in 2002, allowing for the determination of skin corrosion/irritation by applying a tiered testing strategy, using validated *in vitro* or *ex vivo* test methods, thus avoiding pain and suffering of animals. Three validated *in vitro* test methods have been adopted as OECD Test Guidelines 430, 431 and 435 (5)(6)(7), to be used for the corrosivity part of the tiered testing strategy of TG 404 (4).

3. This Test Guideline addresses the human health endpoint skin irritation. It is based on reconstructed human epidermis (RhE), which in its overall design (the use of human derived non-transformed epidermis keratinocytes as cell source and use of representative tissue and cytoarchitecture) closely mimics the biochemical and physiological properties of the upper parts of the human skin, *i.e.* the epidermis. This Test Guideline also includes a set of Performance Standards (PS)(Annex 2) for the assessment of similar and modified RhE-based test methods developed by EC-ECVAM (8), in accordance with the principles of Guidance Document No. 34 (9).

4. Prevalidation, optimisation and validation studies have been completed for an *in vitro* test method (10)(11)(12)(13)(14)(15)(16)(17)(18)(19)(20), using an RhE model, commercially available as EpiSkin™ (designated the Validated Reference Method – VRM). Two other commercially available *in vitro* skin irritation RhE test methods have shown similar results to the VRM according to PS-based validation, and these are the EpiDerm™ SIT (EPI-200) and the SkinEthic™ RHE test methods (21)(22).

5. Before a proposed similar or modified *in vitro* RhE test method for skin irritation can be used for regulatory purposes, its reliability, relevance (accuracy), and limitations for its proposed use should be determined in order to ensure that it can be regarded as similar to that of the VRM, in accordance with the requirements of the PS set out in this Test Guideline (Annex 2).

6. Definitions used are provided in Annex 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

7. A limitation of the Test Guideline, as demonstrated by the validation study (16), is that it only allows classification of chemicals (substances and mixtures) as skin irritants according to UN GHS Category 2, and it does not allow the classification of chemicals to the optional UN GHS Category 3 (mild irritants)(1). However, depending on the regulatory framework in member countries, all non-Category 2 chemicals may be considered as non-irritant (UN GHS No Category). Thus, the regulatory framework in member countries will decide if this Test Guideline will be used as a skin irritation stand-alone replacement test, or as a partial replacement test, within a tiered testing strategy (4). When used as a partial replacement test, follow-up *in vivo* testing may be required to fully characterize skin irritation potential (4). It is recognized that the use of human skin is subject to national and international ethical considerations and conditions.

9. 8. This Test Guideline addresses the *in vitro* skin irritation component of the tiered testing strategy of TG 404 on dermal corrosion/irritation (4). While this Test Guideline does not provide adequate information on skin corrosion, it should be noted that OECD TG 431 on skin corrosion is based on the same RhE test system, though using another protocol (6). This Test Guideline is based on RhE-models using human keratinocytes and therefore addresses *in vitro* the target organ of the species of interest. It moreover directly covers the initial step of the inflammatory cascade/mechanism of action (cell damage and tissue damage resulting in localised trauma) that occurs during irritation *in vivo*. A wide range of chemicals has been tested in the validation underlying this Test Guideline and the empirical database of the validation study amounted to 58 chemicals in total (16)(18)(23). The Test Guideline is applicable to solids, liquids, semi-solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Whenever possible, solids should be ground to a fine powder before application; no other pre-treatment of the sample is required. Gases and aerosols have not been assessed yet in a validation study (24). While it is conceivable that these can be tested using RhE technology, the current Test Guideline does not allow testing of gases and aerosols. It should also be noted that highly coloured chemicals may interfere with the cell viability measurements and need the use of adapted controls for corrections (see paragraphs 24-26).

10. A single testing run should be sufficient when the classification is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements and/or mean percent viability equal to 50 ± 5%, a second run should be considered, as well as a third one in case of discordant results between the first two runs.

PRINCIPLE OF THE TEST

10. The test chemical is applied topically to a three-dimensional RhE model, comprised of non-transformed human-derived epidermal keratinocytes, which have been cultured to form a multilayered, highly differentiated model of the human *epidermis*. It consists of organized basal, spinous and granular layers, and a multilayered *stratum corneum* containing intercellular lamellar lipid layers representing main lipid classes analogous to those found *in vivo*.

11. Chemical-induced skin irritation, manifested by erythema and oedema, is the result of a cascade of events beginning with penetration of the *stratum corneum* and damage to the underlying layers of keratinocytes. The dying keratinocytes release mediators that begin the inflammatory cascade which acts on the cells in the *dermis*, particularly the stromal and endothelial cells. It is the dilation and increased permeability of the endothelial cells that produce the observed erythema and oedema (23). The RhE-based test methods measure the initiating events in the cascade.

12. Cell viability in RhE models is measured by enzymatic conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue; CAS number 298-93-1], into a

blue formazan salt that is quantitatively measured after extraction from tissues (25). Irritant chemicals are identified by their ability to decrease cell viability below defined threshold levels (*i.e.* ≤ 50%, for UN GHS Category 2). Depending on the regulatory framework and applicability of the Test Guideline, chemicals that produce cell viabilities above the defined threshold level, may be considered non-irritants (*i.e.* > 50%, No Category).

DEMONSTRATION OF PROFICIENCY

13. Prior to routine use of any of the three validated test methods that adhere to this Test Guideline, laboratories should demonstrate technical proficiency, using the ten Proficiency Chemicals listed in Table 1. For similar test methods developed under this Test Guideline or for modifications of any of the three validated test methods, the PS requirements described in Annex 2 of this Test Guideline should be met prior to using the test method for regulatory testing.

14. As part of the proficiency exercise, it is recommended that the user verifies the barrier properties of the tissues after receipt as specified by the RhE model producer. This is particularly important if tissues are shipped over long distance/time periods. Once a test method has been successfully established and proficiency in its use has been demonstrated, such verification will not be necessary on a routine basis. However, when using a test method routinely, it is recommended to continue to assess the barrier properties in regular intervals.

Table 1. Proficiency Chemicals[1]

Chemical	CAS NR	*In vivo* score[2]	Physical state	UN GHS Category
naphthalene acetic acid	86-87-3	0	Solid	No Cat.
isopropanol	67-63-0	0.3	Liquid	No Cat.
methyl stearate	112-61-8	1	Solid	No Cat.
heptyl butyrate	5870-93-9	1.7	Liquid	No Cat. (*Optional Cat. 3*)[3]
hexyl salicylate	6259-76-3	2	Liquid	No Cat. (*Optional Cat. 3*)[3]
cyclamen aldehyde	103-95-7	2.3	Liquid	Cat. 2
1-bromohexane	111-25-1	2.7	Liquid	Cat. 2
potassium hydroxide (5% aq.)	1310-58-3	3	Liquid	Cat. 2
1-methyl-3-phenyl-1-piperazine	5271-27-2	3.3	Solid	Cat. 2
heptanal	111-71-7	3.4	Liquid	Cat. 2

[1] The Proficiency Chemicals are a subset of the chemicals used in the validation study.

[2] *In vivo* score in accordance with the OECD Test Guideline 404 (4).

[3] Under this Test Guideline, the UN GHS optional Category 3 (mild irritants)(1) is considered as No Category.

PROCEDURE

15. The following is a description of the components and procedures of a RhE test method for skin irritation assessment. A RhE model should be reconstructed, and can be in-house-prepared or obtained commercially. Standard Operating Procedures (SOPs) for the EpiSkin™, EpiDerm™ and SkinEthic™ RHE are available (26)(27)(28). Testing should be performed according to the following:

RHE TEST METHOD COMPONENTS

General conditions

16. Non-transformed human keratinocytes should be used to reconstruct the epithelium. Multiple layers of viable epithelial cells (basal layer, *stratum spinosum, stratum granulosum*) should be present under a functional *stratum corneum*. *Stratum corneum* should be multilayered containing the essential lipid profile to produce a functional barrier with robustness to resist rapid penetration of cytotoxic marker chemicals, *e.g.* sodium dodecyl sulphate (SDS) or Triton X-100. The barrier function should be demonstrated and may be assessed either by determination of the concentration at which a marker chemical reduces the viability of the tissues by 50% (IC_{50}) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50% (ET_{50}) upon application of the marker chemical at a specified, fixed concentration. The containment properties of the RhE model should prevent the passage of material around the *stratum corneum* to the viable tissue, which would lead to poor modelling of skin exposure. The RhE model should be free of contamination by bacteria, viruses, mycoplasma, or fungi.

Functional conditions

Viability

17. The assay used for determining the magnitude of viability is the MTT-assay (25). The RhE model users should ensure that each batch of the RhE model used meets defined criteria for the negative control (NC). The optical density (OD) of the extraction solvent alone should be sufficiently small, *i.e.* OD<0.1. An acceptability range (upper and lower limit) for the negative control OD values (In the Skin Irritation Test Method conditions) are established by the RhE model developer/supplier, and the acceptability ranges for the 3 validated test methods are given in Table 2. It should be documented that the tissues treated with NC are stable in culture (provide similar viability measurements) for the duration of the test exposure period.

Table 2. Acceptability ranges for negative control OD values

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM)	≥ 0.6	≤ 1.5
EpiDerm™ (EPI-200 SIT)	≥ 1.0	≤ 2.5
SkinEthic™ RHE	≥ 1.2	≤ 2.5

Barrier function

18. The *stratum corneum* and its lipid composition should be sufficient to resist the rapid penetration of cytotoxic marker chemicals, *e.g.* SDS or Triton X-100, as estimated by IC_{50} or ET_{50} (Table 3).

Morphology

19. Histological examination of the RhE model should be performed demonstrating human *epidermis*-like structure (including multilayered *stratum corneum*).

Reproducibility

20. The results of the positive and negative controls of the test method should demonstrate reproducibility over time.

Quality control (QC)

21. The RhE model developer/supplier should ensure and demonstrate that each batch of the RhE model used meets defined production release criteria, among which those for *viability* (paragraph 17), *barrier function* (paragraph 18) and *morphology* (paragraph 19) are the most relevant. These data should be provided to the test method users, so that they are able to include this information in the test report. An acceptability range (upper and lower limit) for the IC_{50} or the ET_{50} should be established by the RhE model developer/supplier (or investigator when using an in-house model). Only results produced with qualified tissues can be accepted for reliable prediction of irritation classification. As an example, the acceptability ranges for the three validated test methods are given in Table 3.

Table 3. Examples of QC batch release criteria

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM) (18 hours treatment with SDS)(26)	IC_{50} = 1.0 mg/ml	IC_{50} = 3.0 mg/ml
EpiDerm™ (EPI-200 SIT) (1% Triton X-100)(27)	ET_{50} = 4.8 hr	ET_{50} = 8.7 hr
SkinEthic™ RHE (1% Triton X-100)(28)	ET_{50} = 4.0 hr	ET_{50} = 9.0 hr

Application of the Test and Control Chemicals

22. At least three replicates should be used for each test substance and for the controls in each run. For liquid as well as solid chemicals, sufficient amount of test chemical should be applied to uniformly cover the *epidermis* surface while avoiding an infinite dose, *i.e.* a minimum of 25 µL/cm^2 or 25 mg/cm^2 should be used. For solid chemicals, the *epidermis* surface should be moistened with deionised or distilled water before application, to improve contact between the test chemical and the *epidermis* surface. Whenever possible, solids should be tested as a fine powder. At the end of the exposure period, the test chemical should be carefully washed from the *epidermis* surface with aqueous buffer, or 0.9% NaCl. Depending on which of the three validated RhE test methods is used, the exposure period varies between 15 and 60 minutes, and the incubation temperature between 20 and 37°C. These exposure periods and temperatures are optimized for each RhE test method and represent the different intrinsic properties of the test methods, for details, see the Standard Operating Protocols (SOPs) for the test methods (26)(27)(28).

23. Concurrent NC and positive controls (PC) should be used in each run to demonstrate that viability (with the NC), barrier function and resulting tissue sensitivity (with the PC) of the tissues are

within a defined historical acceptance range. The suggested PC chemical is 5% aqueous SDS. The suggested NC chemicals are water or phosphate buffered saline (PBS).

Cell Viability Measurements

24. The most important element of the test procedure is that viability measurements are not performed immediately after the exposure to the test chemicals, but after a sufficiently long post-treatment incubation period of the rinsed tissues in fresh medium. This period allows both for recovery from weak cytotoxic effects and for appearance of clear cytotoxic effects. The test optimisation phase (11)(12)(13)(14)(15) demonstrated that a 42 hours post-treatment incubation period was optimal.

25. The MTT assay is a validated quantitative method which should be used to measure cell viability under this Test Guideline. It is compatible with use in a three-dimensional tissue construct. The tissue sample is placed in MTT solution of appropriate concentration (*e.g.* 0.3 - 1 mg/mL) for 3 hours. The precipitated blue formazan product is then extracted from the tissue using a solvent (*e.g.* isopropanol, acidic isopropanol), and the concentration of formazan is measured by determining the OD at 570 nm using a filter band pass of maximum ± 30 nm.

26. Optical properties of the test chemical or its chemical action on the MTT may interfere with the assay leading to a false estimate of viability (because the test chemical may prevent or reverse the colour generation as well as cause it). This may occur when a specific test chemical is not completely removed from the tissue by rinsing or when it penetrates the *epidermis*. If a the test chemical that acts directly on the MTT is naturally coloured, or becomes coloured during tissue treatment, additional controls should be used to detect and correct for test chemical interference with the viability measurement technique. Detailed description of how to correct direct MTT reduction and interferences by colouring agents is available in the SOPs for the three validated test methods (26)(27)(28).

Acceptability Criteria

27. For each test method using valid RhE model batches (see paragraph 21), tissues treated with the NC should exhibit OD reflecting the quality of the tissues that followed shipment, receipt steps and all protocol processes. Control OD values should not be below historically established boundaries. Similarly, tissues treated with the PC, *i.e.* 5% aqueous SDS, should reflect their ability to respond to an irritant chemical under the conditions of the test method (26)(27)(28). Associated and appropriate measures of variability between tissue replicates should be defined (*e.g.* if standard deviations (SD) are used they should be within the 95% viability confidence interval; SD $\leq$ 18%).

Interpretation of Results and Prediction Model

28. The OD values obtained with each test substance can be used to calculate the percentage of viability normalised to NC, which is set to 100%. The cut-off value of percentage cell viability distinguishing irritant from non-classified test chemicals and the statistical procedure(s) used to evaluate the results and identify irritant chemicals, should be clearly defined, documented, and proven to be appropriate. The cut-off values for the prediction of irritation are given below:

The test chemical is considered to be irritant to skin in accordance with UN GHS Category 2 if the tissue viability after exposure and post-treatment incubation is less than or equal (≤) to 50%.

Depending on the regulatory framework in member countries, the test chemical may be considered as non-irritant to skin in accordance with UN GHS No Category if the tissue viability

after exposure and post-treatment incubation is more than (>) 50%.

DATA AND REPORTING

Data

29. For each run, data from individual replicate tissues (*e.g.* OD values and calculated percentage cell viability data for each test chemical, including classification) should be reported in tabular form, including data from repeat experiments as appropriate. In addition means ± SD for each run should be reported. Observed interactions with MTT reagent and coloured test chemicals should be reported for each tested chemical.

Test Report

30. The test report should include the following information:

Test and Control Chemicals:

- Chemical name(s) such as CAS name and number, if known;
- Purity and composition of the chemical (in percentage(s) by weight);
- Physical-chemical properties relevant to the conduct of the study (*e.g.* physical state, stability, volatility, pH and water solubility if known);
- Treatment of the test/control chemicals prior to testing, if applicable (*e.g.* warming, grinding);
- Storage conditions;

Justification of the RhE model and protocol used

Test Conditions:

- Cell system used;
- Complete supporting information for the specific RhE model used including its performance. This should include, but is not limited to;
  - i) Viability
  - ii) Barrier function
  - iii) Morphology
  - iv) Reproducibility and predictivity
  - v) Quality controls (QC) of the model
- Details of the test procedure used;
- Test doses used, duration of exposure and post treatment incubation period;
- Description of any modifications of the test procedure;
- Reference to historical data of the model. This should include, but is not limited to;
  - i) acceptability of the QC data with reference to historical batch data
  - ii) acceptability of the positive and negative control values with reference to positive and negative control means and ranges
- Description of evaluation criteria used including the justification for the selection of the cut-off point(s) for the prediction model;
- Reference to historical control data;

Results:

- Tabulation of data from individual test substances for each run;
- Description of other effects observed;

Discussion of the results

Conclusion

LITERATURE

1) UN (2009), United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Third revised edition, UN New York and Geneva. Available at: [http://www.unece.org/trans/danger/publi/ghs/ghs_rev03/03files_e.html]
2) EC-ECVAM (2009), Statement on the "Performance under UN GHS of three *in vitro* assays for skin irritation testing and the adaptation of the Reference Chemicals and Defined Accuracy Values of the ECVAM skin irritation Performance Standards", issued by the ECVAM Scientific Advisory Committee (ESAC30), 9 April 2009. Available under "*Publications*" at: [http://ecvam.jrc.ec.europa.eu]
3) EC (2008), REGULATION (EC) No 1272/2008 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Official Journal of the European Union L353, 1-1355.
4) OECD (2004), Acute dermal irritation/corrosion. OECD Guideline for the Testing of Chemicals No. 404, OECD, Paris. . Available at: [http://www.oecd.org/env/testguidelines]
5) OECD (2004), *In vitro* skin corrosion: Transcutaneous Electrical Resistance (TER). OECD Guideline for the Testing of Chemicals No. 430, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]
6) OECD (2004), *In vitro* skin model. OECD Guideline for the Testing of Chemicals No. 431, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]
7) OECD (2006), *In vitro* membrane barrier test method. OECD Guideline for the Testing of Chemicals No. 435, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines
8) EC-ECVAM (2009), Performance Standards for *in vitro* skin irritation test methods based on Reconstructed human Epidermis (RhE). Available under "Validation Study Documents, section Skin Irritation", at: [http://ecvam.jrc.ec.europa.eu]
9) OECD (2005), Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. OECD Series on Testing and Assessment No. 34, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]
10) Fentem, J. H., Briggs, D., Chesné, C., Elliot, G. R., Harbell, J. W., Heylings, J. R., Portes, P., Roguet, R., van de Sandt, J. J. M. and Botham, P. (2001), A prevalidation study on *in vitro* tests for acute skin irritation. Results and evaluation by the Management Team. *Toxicol. in Vitro* 15, 57-93.
11) Portes, P., Grandidier, M.-H., Cohen, C. and Roguet, R. (2002), Refinement of the EPISKIN protocol for the assessment of acute skin irritation of chemicals: follow-up to the ECVAM prevalidation study. *Toxicol. in Vitro* 16, 765–770.
12) Kandárová, H., Liebsch, M., Genschow, E., Gerner, I., Traue, D., Slawik, B. and Spielmann, H. (2004), Optimisation of the EpiDerm test protocol for the upcoming ECVAM validation study on *in vitro* skin irritation tests. *ALTEX* 21, 107–114.
13) Kandárová, H., Liebsch, M., Gerner, I., Schmidt, E., Genschow, E., Traue, D. and Spielmann, H. (2005), The EpiDerm test protocol for the upcoming ECVAM validation study on *in vitro* skin irritation tests – An assessment of the performance of the optimised test. *ATLA* 33, 351-367.
14) Cotovio, J., Grandidier, M.-H., Portes, P., Roguet, R. and Rubinsteen, G. (2005), The *in vitro* acute skin irritation of chemicals: optimisation of the EPISKIN prediction model within the framework of the ECVAM validation process. *ATLA* 33, 329-349.
15) Zuang, V., Balls, M., Botham, P. A., Coquette, A., Corsini, E., Curren, R. D., Elliot, G. R., Fentem, J. H., Heylings, J. R., Liebsch, M., Medina, J., Roguet, R., van De Sandt, J. J. M., Wiemann, C. and Worth, A. (2002), Follow-up to the ECVAM prevalidation study on *in vitro* tests

for acute skin irritation. The European Centre for the Validation of Alternative Methods Skin Irritation Task Force report 2. *ATLA* 30, 109-129.

16) Spielmann, H., Hoffmann, S., Liebsch, M., Botham, P., Fentem, J., Eskes, C., Roguet, R., Cotović, J., Cole, T., Worth, A., Heylings, J., Jones, P., Robles, C., Kandárová, H., Gamer, A., Remmele, M., Curren, R., Raabe, H., Cockshott, A., Gerner, I. and Zuang, V. (2007), The ECVAM international validation study on *in vitro* tests for acute skin irritation: Report on the validity of the EPISKIN and EpiDerm assays and on the skin integrity function test. *ATLA* 35, 559-601.

17) Hoffmann, S. (2006), ECVAM skin irritation validation study phase II: Analysis of the primary endpoint MTT and the secondary endpoint IL1-α. Available under "*Validation Study Documents, section Skin Irritation*", at: [http://ecvam.jrc.ec.europa.eu]

18) Eskes, C., Cole, T., Hoffmann, S., Worth, A., Cockshott, A., Gerner, I. and Zuang, V. (2007), The ECVAM international validation study on *in vitro* tests for acute skin irritation: selection of test chemicals. *ATLA* 35, 603-619.

19) Cotovio, J., Grandidier, M.-H., Lelièvre, D., Roguet, R., Tinois-Tessonneaud, E. and Leclaire, J. (2007), *In vitro* acute skin irritancy of chemicals using the validated EPISKIN model in a tiered strategy - Results and performances with 184 cosmetic ingredients, *AATEX*, Special Issue-proceedings from WC6. Vol.14, 351-358.

20) EC-ECVAM (2007), Statement on the validity of in-vitro tests for skin irritation, issued by the ECVAM Scientific Advisory Committee (ESAC26), 27 April 2007. Available under "Publications" at: [http://ecvam.jrc.ec.europa.eu]

21) EC-ECVAM (2007), Performance Standards for applying human skin models to *in vitro* skin irritation testing. Available under "*Validation Study Documents, section Skin Irritation*", at: [http://ecvam.jrc.ec.europa.eu]

22) EC-ECVAM (2008), Statement on the scientific validity of *in vitro* tests for skin irritation testing, issued by the ECVAM Scientific Advisory Committee (ESAC29), 5 November 2008. Available under "*Publications*"at: [http://ecvam.jrc.ec.europa.eu]

23) OECD (20xx), Explanatory background document to the OECD draft test guideline on *in vitro* skin irritation testing. To be published in OECD Series on Testing and Assessment, No. 1XX, OECD, Paris. Available at:

24) Welss, T., Basketter, D. A. and Schröder, K. R. (2004), *In vitro* skin irritation: fact and future. State of the art review of mechanisms and models. *Toxicol. in Vitro* 18, 231-243.

25) Mosmann, T. (1983), Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods* 65, 55-63.

26) EpiSkin™ SOP, Version 1.8 (February 2009), ECVAM Skin Irritation Validation Study: Validation of the EpiSkin™ test method $^{15\text{ min - }42\text{ hours}}$ for the prediction of acute skin irritation of chemicals. Available under "*Validation Study Documents, section Skin Irritation*", at: [http://ecvam.jrc.ec.europa.eu]

27) EpiDerm™ SOP, Version 7.0 (Revised March 2009), Protocol for: In vitro EpiDerm™ skin irritation test (EPI-200-SIT). For use with MatTek Corporation's reconstructed human epidermal model EpiDerm (EPI-200). Available under "*Validation Study Documents, section Skin Irritation*", at: [http://ecvam.jrc.ec.europa.eu]

28) SkinEthic™ RHE SOP, Version 2.0 (February 2009), SkinEthic skin irritation test$^{-42\text{bis}}$ test method for the prediction of acute skin irritation of chemicals: 42 minutes application + 42 hours post-incubation. Available under "*Validation Study Documents, section Skin Irritation*", at: [http://ecvam.jrc.ec.europa.eu]

29) Harvell, J. D., Lamminstausta, K., and Maibach, H. I. (1995), Irritant contact dermatitis. In: *Practical Contact Dermatitis*, pp 7-18, (Ed. Guin J. D.). Mc Graw-Hill, New York.

30) EC (2001), Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Official Journal of the European Union* **L225**, 1-333.

31) Basketter, D. A., York, M., McFadden, J. P. and Robinson, M. K. (2004), Determination of skin irritation potential in the human 4-h patch test. *Contact Dermatitis* 51, 1-4.

32) Jirova, D., Liebsch, M., Basketter, D., Spiller, E., Kejlova, K., Bendova, H., Marriott, M. and Kandarova, H. (2007), Comparison of human skin irritation and photo-irritation patch test data with cellular *in vitro* assays and animal *in vivo* data. *AATEX, Special Issue-proceedings from WC6*. Vol. 14, 359-365.

ANNEX 1

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method (9).

Cell viability: Parameter measuring total activity of a cell population *e.g.* as ability of cellular mitochondrial dehydrogenases to reduce the vital dye MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue), which depending on the endpoint measured and the test design used, correlates with the total number and/or vitality of living cells.

Concordance: This is a measure of test method performance for test methods that give a categorical result, and is one aspect of relevance. The term is sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of substances being examined (9).

ET_{50}: Can be estimated by determination of the exposure time required to reduce cell viability by 50% upon application of the marker chemical at a specified, fixed concentration, see also IC_{50}.

EU CLP (European Commission Regulation on the Classification, Labelling and Packaging of Substances and Mixtures): Implements in the European Union (EU) the UN GHS system for the classification of chemicals (substances and mixtures)(3).

GHS (Globally Harmonized System of Classification and Labelling of Chemicals by the United Nations (UN)): A system proposing the classification of chemicals (substances and mixtures) according to standardized types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (1).

IC_{50}: Can be estimated by determination of the concentration at which a marker chemical reduces the viability of the tissues by 50% (IC_{50}) after a fixed exposure time, see also ET_{50}.

Infinite dose: Amount of test chemical applied to the *epidermis* exceeding the amount required to completely and uniformly cover the *epidermis* surface.

Me-too test: A colloquial expression for a test method that is structurally and functionally similar to a validated and accepted reference test method. Such a test method would be a candidate for catch-up validation. Interchangeably used with similar test method (9).

Mixture: Used in the context of the UN GHS (1) as a mixture or solution composed of two or more substances in which they do not react.

Performance standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are; (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method;

and (iii) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (9).

Reference chemicals: Chemicals selected for use in the validation process, for which responses in the *in vitro* or *in vivo* reference test system or the species of interest are already known. These chemicals should be representative of the classes of chemicals for which the test method is expected to be used, and should represent the full range of responses that may be expected from the chemicals for which it may be used, from strong, to weak, to negative. Different sets of reference chemicals may be required for the different stages of the validation process, and for different test methods and test uses (9).

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (9).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability (9).

Replacement test: A test which is designed to substitute for a test that is in routine use and accepted for hazard identification and/or risk assessment, and which has been determined to provide equivalent or improved protection of human or animal health or the environment, as applicable, compared to the accepted test, for all possible testing situations and substances (9).

Sensitivity: The proportion of all positive/active substances that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (9).

Skin irritation: The production of reversible damage to the skin following the application of a test chemical for up to 4 hours. Skin irritation is a locally arising, non-immunogenic reaction, which appears shortly after stimulation (29). Its main characteristic is its reversible process involving inflammatory reactions and most of the clinical characteristic signs of irritation (erythema, oedema, itching and pain) related to an inflammatory process.

Specificity: The proportion of all negative/inactive substances that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (9).

Substance: Used in the context of the UN GHS (1) chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the stability of the product and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Tiered testing strategy: Testing which uses test methods in a sequential manner; the test methods selected in each succeeding level are determined by the results in the previous level of testing (9).

ANNEX 2

PERFORMANCE STANDARDS FOR ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED *IN VITRO* RECONSTRUCTED HUMAN *EPIDERMIS* (RhE) TEST METHODS FOR SKIN IRRITATION

INTRODUCTION

1. The purpose of Performance Standards (PS) is to communicate the basis by which new test methods, both proprietary (*i.e.* copyrighted, trademarked, registered) and non-proprietary can be determined to have sufficient accuracy and reliability for specific testing purposes. These PS, based on validated and accepted test methods, can be used to evaluate the reliability and accuracy of other analogous test methods (colloquially referred to as "me-too" tests) that are based on similar scientific principles and measure or predict the same biological or toxic effect (9).

2. Prior to adoption of modified test methods, *i.e.* proposed potential improvements to an approved test method, there should be an evaluation to determine the effect of the proposed changes on the test's performance and the extent to which such changes affect the information available for the other components of the validation process. Depending on the number and nature of the proposed changes, the generated data and supporting documentation for those changes, they should either be subjected to the same validation process as described for a new test, or, if appropriate, to a limited assessment of reliability and relevance using established PS (9).

3. Similar (me-too) or modified test methods of any of the three validated test methods [EpiSkin™ (Validated Reference Method – VRM), EpiDerm™ SIT and SkinEthic™ RHE] proposed for use under this Test Guideline should be evaluated to determine their reliability and accuracy using chemicals representing the full range of the Draize irritancy scores. When evaluated using the 20 recommended Reference Chemicals of the PS (Table 1), the proposed similar or modified test methods should have reliability and accuracy values which are comparable or better to those derived from the VRM (Table 2) (2)(16). The reliability and accuracy values that should be achieved are provided in paragraphs 8 to 12 of this Annex. Non-classified (UN GHS No Category) and classified (UN GHS Category 2)(1) chemicals, representing different chemical classes are included, so that the reliability and accuracy (sensitivity, specificity and overall accuracy) of the proposed test method can be compared to that of the VRM. The reliability of the test method, as well as its ability to correctly identify UN GHS Category 2 irritant chemicals and, depending on the regulatory framework in member countries, also its ability to correctly identify UN GHS No Category chemicals (for member countries that do not adopt optional UN GHS Category 3), should be determined prior to its use for testing new substances.

4. These PS are based on the EC-ECVAM PS (8), updated according to the UN GHS and EU CLP systems on classification and labelling (1)(23). The original PS were defined after the completion of the validation study (21) and were based on the EU classification system as described in the 28th amendment to the Dangerous Substances Directive (30). Due to the adoption of the UN GHS system for classification and labelling in EU (EU CLP)(23), which took place between the finalisation of the validation study and the completion of this Test Guideline, the PS have been updated (8). This update concerns mainly changes; (i), in the set of the PS Reference Chemicals; and (ii), the defined reliability and accuracy values (2)(23).

PERFORMANCE STANDARDS FOR *IN VITRO* RhE TEST METHODS FOR SKIN IRRITATION

5. The PS comprises the following three elements (9):

I) Essential Test Method Components

II) Minimum List of Reference Chemicals

III) Defined Reliability and Accuracy Values

I) Essential Test Method Components

6. These consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed, mechanistically and functionally similar or modified test method. These components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components will help to assure that a similar or modified proposed test method is based on the same concepts as the corresponding VRM (9). The essential test method components are described in detail in paragraphs 16 to 21 of the Test Guideline and testing should be performed according to the following:

The general conditions (paragraph 16)

The functional conditions, which include:

viability (paragraph 17);

barrier function (paragraph 18);

morphology (paragraph 19);

reproducibility (paragraph 20); and,

quality control (paragraph 21)

II) Minimum List of Reference Chemicals

7. Reference Chemicals are used to determine if the reliability and accuracy of a proposed similar or modified test method, proven to be structurally and functionally sufficiently similar to the VRM, or representing a minor modification of one of the three validated test methods, are comparable or better than those of the VRM (2)(8)(16)(23). The 20 recommended Reference Chemicals listed in Table 1 include chemicals representing different chemical classes, and are representative of the full range of Draize irritancy scores (from non-irritant to strong irritant). The chemicals included in this list comprise 10 UN GHS Category 2 chemicals and 10 non-categorised chemicals, of which 3 are optional UN GHS Category 3 chemicals. Under this Test Guideline, the optional Category 3 is considered as No Category. The chemicals listed in Table 1 are selected from the chemicals used in the optimisation phase that followed prevalidation and in the validation study of the VRM, with regard to chemical functionality and physical state (14)(18). These Reference Chemicals represent the minimum number of chemicals that should be used to evaluate the accuracy and reliability of a proposed similar or modified test method, but should not be used for development and/or optimisation of new test methods. In situations where a listed chemical is unavailable, other chemicals for which adequate *in vivo* reference data are available could be used, primarily from the chemicals used in the optimisation phase following prevalidation or the validation study of the VRM. If desired, additional chemicals representing other chemical classes and for which adequate *in vivo* reference data are available may be added to the minimum list of Reference Chemicals to further evaluate the accuracy of the proposed test method.

Table 1. Minimum List of Reference Chemicals for determination of Accuracy and Reliability Values for Similar or Modified RhE Skin Irritation Test Methods[1]

Chemical	CAS Number	Physical state	*In vivo* score	VRM *in vitro* Cat.	UN GHS *in vivo* Cat.
1-bromo-4-chlorobutane	6940-78-9	Liquid	0	Cat. 2	No Cat.
diethyl phthalate	84-66-2	Liquid	0	No Cat.	No Cat.
naphthalene acetic acid	86-87-3	Solid	0	No Cat.	No Cat.
allyl phenoxy-acetate	7493-74-5	Liquid	0.3	No Cat.	No Cat.
isopropanol	67-63-0	Liquid	0.3	No Cat.	No Cat.
4-methyl-thio-benzaldehyde	3446-89-7	Liquid	1	Cat. 2	No Cat.
methyl stearate	112-61-8	Solid	1	No Cat.	No Cat.
heptyl butyrate	5870-93-9	Liquid	1.7	No Cat.	No Cat. (*Optional Cat. 3*)
hexyl salicylate	6259-76-3	Liquid	2	No Cat.	No Cat. (*Optional Cat. 3*)
cinnamaldehyde	104-55-2	Liquid	2	Cat. 2	No Cat. (*Optional Cat. 3*)
1-decanol[2]	*112-30-1*	*Liquid*	*2.3*	*Cat. 2*	*Cat. 2*
cyclamen aldehyde	103-95-7	Liquid	2.3	Cat. 2	Cat. 2
1-bromohexane	111-25-1	Liquid	2.7	Cat. 2	Cat. 2
2-chloromethyl-3,5-dimethyl-4-methoxypyridine HCl	86604-75-3	Solid	2.7	Cat. 2	Cat. 2
di-n-propyl disulphide[2]	*629-19-6*	*Liquid*	*3*	*No Cat.*	*Cat. 2*
potassium hydroxide (5% aq.)	1310-58-3	Liquid	3	Cat. 2	Cat. 2
benzenethiol, 5-(1,1-dimethylethyl)-2-methyl	7340-90-1	Liquid	3.3	Cat. 2	Cat. 2
1-methyl-3-phenyl-1-piperazine	5271-27-2	Solid	3.3	Cat. 2	Cat. 2
heptanal	111-71-7	Liquid	3.4	Cat. 2	Cat. 2
tetrachloroethylene	127-18-4	Liquid	4	Cat. 2	Cat. 2

[1] The chemical selection is based on the following criteria; (i), the chemicals are commercially available; (ii), they are representative of the full range of Draize irritancy scores (from non-irritant to strong irritant); (iii), they have a well-defined chemical structure; (iv), they are representative of the chemical functionality used in the validation process; and (v), they are not associated with an extremely toxic profile (*e.g.* carcinogenic or toxic to the reproductive system) and they are not associated with prohibitive disposal costs.

[2] Chemicals that are irritant in the rabbit but for which there is reliable evidence that they are non-irritant in humans (31)(32).

III) Defined Reliability and Accuracy Values

8. For purposes of small scale validation studies to establish the reliability and relevance of proposed similar or modified test methods to be transferred between laboratories, all 20 Reference Chemicals should be tested in at least three laboratories. However, if the proposed test method is to be used in a single laboratory only, multi laboratory testing will not be required for validation. It is however essential that such validation studies are independently assessed by internationally recognised validation bodies, in agreement with international guidelines (9). In each laboratory, all 20 Reference Chemicals should be tested in three independent runs performed with different tissue batches and at sufficiently spaced time points. Each run should consist of a minimum of three concurrently tested tissue replicates for each included test substance, NC and PC. However, the number of concurrently tested tissue replicates recommended per run (all cases) may be reduced if statistically/scientifically sufficiently justified (*e.g.* at least 90% of concurrently treated tissue replicates during development/optimization of the test method showing a difference of viability below 5%).

9. The calculation of the reliability and accuracy values of the proposed test method should be based on considering all four criteria below together, ensuring that the values for reliability and relevance are calculated in a predefined and consistent manner:

1. Only the data of runs from complete run sequences qualify for the calculation of the test method within, and between-laboratory variability and predictive capacity (accuracy).

2. The final classification for each Reference Chemicals in each participating laboratory should be obtained by using the mean value of viability over the different runs of a complete run sequence.

3. Only the data obtained for chemicals that have complete run sequences in all participating laboratories qualify for the calculation of the test method between-laboratory variability.

4. The calculation of the accuracy values should be done on the basis of the individual laboratory predictions obtained for the 20 Reference Chemicals by the different participating laboratories.

In this context, a **run sequence** consists of three independent runs from one laboratory for one test substance. A **complete run sequence** is a run sequence from one laboratory for one test substance where all three runs are valid. This means that any single invalid run invalidates an entire run sequence of three runs.

Within-laboratory reproducibility

10. An assessment of within-laboratory variability should show a concordance of classifications (UN GHS Category 2 and No Category) obtained in different, independent test runs of the 20 Reference Chemicals within one single laboratory equal or higher (≥) than 90%.

Between-laboratory reproducibility

11. An assessment of between-laboratory reproducibility is not essential if the proposed test method is to be used in a single laboratory only. For methods to be transferred between laboratories, the concordance of classifications (UN GHS Category 2 and No Category) obtained in different, independent test runs of the 20 Reference Chemicals between preferentially a minimum of three laboratories should be equal or higher (≥) than 80%.

Predictive capacity (accuracy)

12. The accuracy (sensitivity, specificity and overall accuracy) of the proposed similar or modified test method should be comparable or better to that of the VRM, taking into consideration additional information relating to relevance in the species of interest (Table 2). The sensitivity should be equal or higher (≥) than 80% (2)(8)(23). However, a further specific restriction applies to the sensitivity of the proposed *in vitro* test method inasmuch as only two *in vivo* Category 2 substances, 1-decanol and di-n-propyl disulphide, may be misclassified as No Category by more than one participating laboratory. The specificity should be equal or higher (≥) than 70% (2)(8)(23). There is no further restriction with regard to the specificity of the proposed *in vitro* test method, ***i.e.*** any participating laboratory may misclassify any *in vivo* No Category substance as long as the final specificity of the test method is within the acceptable range. The overall accuracy should be equal or higher (≥) than 75% (2)(8)(23). Although the sensitivity of the VRM calculated for the 20 Reference Chemicals listed in Table 1 is equal to 90%, the defined minimum sensitivity value required for any similar or modified test method to be considered valid is set at 80% since both 1-decanol (a borderline chemical) and di-n-propyl disulphide (a false negative of the VRM) are known to be non-irritant in humans (31)(32), although being identified as irritants in the rabbit test. Since RhE models are based on cells of human origin, they may predict these chemicals as non-irritant (UN GHS No Category)

Table 2. Required predictive values for any similar or modified test method to be considered valid.

Sensitivity	Specificity	Overall Accuracy
80%	70%	75%

Study Acceptance Criteria

13. It is possible that one or several tests pertaining to one or more test substances does/do not meet the test acceptance criteria for the test and control substances or is/are not acceptable for other reasons. To complement missing data, for each test substance a maximum number of two additional tests is admissible ("retesting"). More precisely, since in case of retesting also PC and NC have to be concurrently tested, a maximum number of two additional runs may be conducted for each test substance.

14. It is conceivable that even after retesting, the minimum number of three valid runs required for each tested substance is not obtained for every Reference Chemicals in every participating laboratory, leading to an incomplete data matrix. In such cases the following apply, taking into account all three criteria below together, with regard to the acceptability of datasets for purposes of similar validation studies:

1. All 20 Reference Chemicals should have at least one complete run sequence.

2. In each of at least three participating laboratories, a minimum of 85% of the run sequences need to be complete (for 20 chemicals: 3 invalid run sequences allowed in a single laboratory).

3. A minimum of 90% of all possible run sequences from at least three laboratories need to be complete (for 20 chemicals tested in 3 laboratories: 6 invalid run sequences allowed in total).

In this context, a **run sequence** consists of three independent runs from one laboratory for one test substance. A **complete run sequence** is a run sequence from one laboratory for one test substance where

all three runs are valid. This means that any single invalid run invalidates an entire run sequence of three runs.